

5SRNA 二级结构的研究

1. 一种以Fox 模型为基础的微型计算机方法

刘次全 王莹 周明培 李靖炎

(中国科学院昆明动物研究所)

摘 要

本文系以 Fox 模型为兰本而设计的一种建立5S RNA二级结构的微型计算机方法。将5S RNA 分子的核苷酸顺序作为信息输入, 经 Z-80微型计算机处理约 4 分钟后, 即可建立在仅考虑二级相互作用的情况下, 相当于 Fox 模型的具有最低自由能的二级结构。

本方法获得的结果与同样以 Fox 模型为兰本的其它方法所得到的结果相符。本方法简便、快速和费用低廉, 易于推广应用。

迄今为止, 有关 5SRNA 二级结构的研究, 尽管运用了物理方法 (如分析离心、紫外吸收、园二色谱、旋光色散、拉曼光谱、小角度X射线衍射和核磁共振等)、化学修饰法、部分酶解法和寡核苷酸结合法等等, 但还没有得到一致公认的二级结构模型。研究表明, 由于采用的研究方法不同, 即使是同一种5S RNA 分子, 也会得出不同的二级结构图 (Erdman等, 1976)。近几年来, 在一些学者提出的模型中, Fox 模型与大多数物理、化学方法测得的结果是相符的。就大肠杆菌5S RNA 而言, 该模型与核磁共振分析存在有28对碱基对的结果是一致的 (李楠茜等, 1982; Fox 等, 1975)。

为探讨 5SRNA二级结构的生物学意义, 我们在 Fox 模型的基础上, 设计了这种在 Z-80微型计算机上即可实现的建立5S RNA 二级结构的计算机方法。同时, 我们按照自由能最低原则还提出了另一种模型 (另文发表)。欲得到某种5S RNA 分子的一类 Fox 模型的二级结构图, 只需将已知的该分子的核苷酸顺序作为信息输入, 经计算机处理后即可打印出所需的结果。表 1 即是应用本方法所得到的大肠杆菌 (*Escherichia Coli*) 5SRNA 二级结构的计算机报告。

* 云南省电子计算站荣茂森同志参加了部分工作, 特此致谢。

本文1983年3月10日收到。

方 法

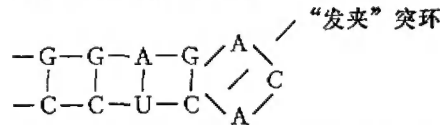
模型处理

我们是以 Fox 等 (Fox 等, 1975) 提出的模型 (图 1 系大肠杆菌 5SRNA 二级结构的 Fox 模型) 为兰本, 按下述规则来处理的:

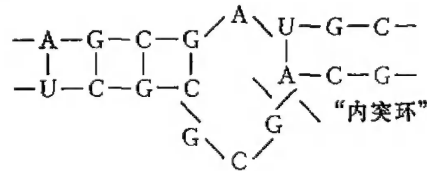
1. 一个螺旋区是由连续的 G—C、A—U 和 G—U 碱基对构成的。一个螺旋区至少需有三个碱基对。

2. 碱基配对是: G 跟 C, G 跟 U 和 A 跟 U 配对。然而, G—U 碱基对不能在一个螺旋区的两端中的任何一端参与形成螺旋区 (此点与 Fox 模型不同)。这是由于考虑到螺旋区的稳定性的缘故。

3. 一个“发夹”突环必需包含至少三个碱基, 其自由能则取决于关闭此“发夹”突环的碱基对 (G—C 或是 A—U)。下图表示由 G—C 对关闭的“发夹”突环:



4. 一个“内突环”的自由能取决于究竟是由两对 G—C 对, 或是一对 G—C 和一对 A—U 对, 还是两对 A—U 对关闭这一突环。下图表示由一对 G—C 和一对 A—U 对关闭的“内突环”:



5. 一个“突起”突环的自由能由横跨这一“突起”的两对碱基对的堆积自由能来确定。必须指出, 在大肠杆菌 5SRNA 二级结构的 Fox 模型中没有“突起”突环。

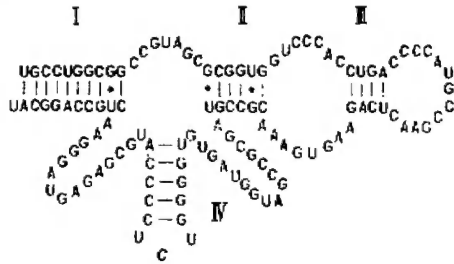
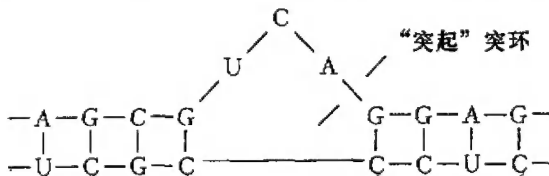
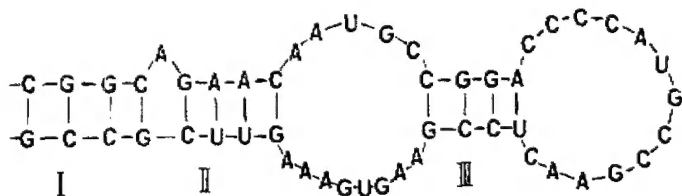


图 1 大肠杆菌 (*E. Coli*) 5SRNA
二级结构模型 (Fox, 1975)

螺旋区(碱基对区)是通过突环连接起来的。下图表示螺旋区 I 和 II 由“突起”突环连接, II 和 III 之间由“内突环”连接:



自由能分布

从热力学知, 一个稳定的 RNA 分子的二级结构应该使溶液的自由能最低。假定 RNA 分子的双链结构对单链结构的自由能记为 ΔG , 则:

$$\Delta G = \Delta H - T \Delta S$$

如果 ΔH 和 ΔS 与绝对温度 T 无关, 于是 RNA 分子内双链形成的总自由能 (ΔG_t) 为碱基对起始形成的自由能 (ΔG_i) 与相继增长的自由能 (ΔG_p) 之和

$$\begin{aligned} \Delta G_t &= \Delta G_i + \Delta G_p \\ &= \Delta H \left(1 - \frac{T}{T_m} \right) - 2.3RT \log \gamma_m \end{aligned}$$

式中 T_m 为熔解温度, γ_m 为 m 个未键合碱基的突环形成的概率, R 为气体常数。

表 2 列出了本文采用的自由能数值* (Tinoco 等, 1973)。

计算机程序

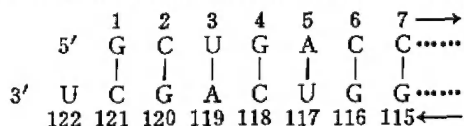
建立一种具有最低自由能的结构, 能量函数是一关键因素。因为它可以排除不可能的或不需要的结构。必须强调指出, 用一种数学处理来选择具有最低自由能的结构, 其处理的量(结构数量)是非常庞大的, 即使是 5SRNA 这样的小分子也是如此。

计算机程序系运用 FORTRAN IV 算法语言编制的两个子程序组合成的。表 1 所列自由能数值编制在计算机程序中。在输入给定的 5SRNA 分子的已知核苷酸顺序, 如:

1 2 3 4 5 6 7
U A G G C G A.....

后, 程序按 G 与 C (或 U)、A 与 U 配对在螺旋区形成氢键, 不能配对的碱基形成突环的原则, 对所有可能的螺旋区进行编辑。

螺旋区由三对或三对以上碱基对顺次排列成反平行的键, 如:



* 自由能数值按文献(Tinoco等, 1973)作了调整和补充。

表 2

25°C RNA二级结构的自由能分布

碱基对区域		ΔG (千卡) $\pm 10\%$
$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{A}}\text{---}\dot{\text{A}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{U}}\text{---}\dot{\text{U}}\text{---} \\ \text{---}\leftarrow \end{array}$	$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{G}}\text{---}\dot{\text{A}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{U}}\text{---}\dot{\text{U}}\text{---} \\ \text{---}\leftarrow \end{array}$	-1.2
$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{A}}\text{---}\dot{\text{U}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{U}}\text{---}\dot{\text{A}}\text{---} \\ \text{---}\leftarrow \end{array}$	$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{U}}\text{---}\dot{\text{A}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{A}}\text{---}\dot{\text{U}}\text{---} \\ \text{---}\leftarrow \end{array}$	-1.8
$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{A}}\text{---}\dot{\text{C}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{U}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \end{array}$	$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{C}}\text{---}\dot{\text{A}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{G}}\text{---}\dot{\text{U}}\text{---} \\ \text{---}\leftarrow \end{array}$	-2.2
$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{A}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{U}}\text{---}\dot{\text{C}}\text{---} \\ \text{---}\leftarrow \end{array}$	$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{A}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{U}}\text{---}\dot{\text{C}}\text{---} \\ \text{---}\leftarrow \end{array}$	-3.2
$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{C}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{G}}\text{---}\dot{\text{C}}\text{---} \\ \text{---}\leftarrow \end{array}$	$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{G}}\text{---}\dot{\text{C}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{C}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \end{array}$	-5.0
$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{C}}\text{---}\dot{\text{C}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{G}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \end{array}$	$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{G}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{C}}\text{---}\dot{\text{C}}\text{---} \\ \text{---}\leftarrow \end{array}$	-0.3
$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{G}}\text{---}\dot{\text{U}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{U}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \end{array}$	$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{U}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{G}}\text{---}\dot{\text{U}}\text{---} \\ \text{---}\leftarrow \end{array}$	-1.3 (内部的)
$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{G}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{U}}\text{---}\dot{\text{C}}\text{---} \\ \text{---}\leftarrow \end{array}$	$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{G}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{C}}\text{---}\dot{\text{U}}\text{---} \\ \text{---}\leftarrow \end{array}$	-2.1 (末端的)
$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{G}}\text{---}\dot{\text{C}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{U}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \end{array}$	$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{C}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{U}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \end{array}$	-1.3 (内部的)
$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{U}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{G}}\text{---}\dot{\text{C}}\text{---} \\ \text{---}\leftarrow \end{array}$	$\begin{array}{c} \text{---}\rightarrow \\ \text{---}\dot{\text{G}}\text{---}\dot{\text{U}}\text{---} \\ \text{---}\leftarrow \\ \text{---}\dot{\text{C}}\text{---}\dot{\text{G}}\text{---} \\ \text{---}\leftarrow \end{array}$	-2.1 (末端的)
未结合区域的未结合碱基数		ΔG (千卡) ± 1 千卡
2—6		“内突环”
7—20		+ 2
$m (>20)$		+ 3
1		+ 1 + 2 log m
2—3		“突起”突环
4—7		+ 3
8—20		+ 4
$m (>20)$		+ 5
3		+ 6
4—5		+ 4 + 2 log m
6—7		“发夹突环”
8—9		G—C关闭 A—U关闭
10—30		+ 8 > 8
$m (>30)$		+ 5 + 7
		+ 4 + 6
		+ 5 + 7
		+ 6 + 8
		3.2 + 2 log m 5.5 + 2 log m

在程序中每个 5S RNA 序列的碱基由顺序编号来表示: 5' 端的第一个碱基编为 1 号, 依次为 2, 3, 4, 5, …… , 直至 3' 端的第一个碱基 (即序列的最后一个碱基)。在编辑时程序作如下处理:

5' 端依次 3 个碱基与 3' 端的 3 个碱基相比较,

```

//      // 4 个 //      //      4 个 //      // ,
//      // 5 个 //      //      5 个 //      // ,
//      //      //      //      //      //
//      //      //      //      //      //
//      // n 个 //      //      n 个 //      // 。

```

找出符合规则 1 和 2 的区域, 然后将两端相匹配的序号分别存入 MX, NX 两个矩阵中, 进而再把这些区域与其相应的自由能联系起来。最后, 将重复的区域以及在能量上和空间上不利的区域除去, 从而找出以 Fox 模型为基础的, 在仅考虑二级相互作用的情况下具有最低总自由能的结构。

本方法除去了螺旋区两端 G—U 碱基对对于螺旋区的自由能贡献 (见规则 2), 并注意在相同碱基重复存在的地方对螺旋结构稳定性的影响。原则上根据螺旋形堆积的负自由能贡献和突环形成的正自由能贡献之和, 就能直接计算以某种方式排列的 5SRNA 二级结构的稳定性。

结 果

根据我们对 49 个已知其核苷酸顺序的 5SRNA 分子和酵母丙氨酸 tRNA 分子二级结构的处理结果表明:

1) 用本文报告的方法建立相应于 Fox 模型的 5SRNA 和“三叶草”型的 tRNA 分子可能的二级结构图是相当理想的。建立的二级结构跟有关文献报告的结果基本相符 (Darlix 等, 1981; Delihis 等, 1981; Fox 等, 1975; Jordan 等, 1974; Luehrsen 等, 1981; Studnicka 等, 1981)。

2) 本方法对小分子 RNA 是普遍适用的, 可用于任何直到 140 个碱基长度的多核苷酸链 (对多于 140 个碱基长度的 RNA 分子, 我们尚未试算过)。

3) 对每一种可能的二级结构同时给出一个总自由能数值, 但处理的结果只打印出具有最低总自由能值的二级结构。

4) 在一般微型计算机上即可进行。本方法就是在 Z—80 微型计算机上建立和通过的。处理一个 5SRNA 分子的二级结构仅需 3~4 分钟, 加上人工画出二级结构图的时间总共也不过 8~10 分钟。

本方法尤其具有简便、快速和费用低廉等特点, 且易于推广应用。

本文所述之方法在建立以 Fox 模型为基础的 5SRNA 可能的二级结构时, 是有一定的实际意义和应用价值的。

参 考 文 献

- 李楠茜, 赵翼、魏西平 1982 5SRNA 的结构, 生物化学与生物物理进展 (6): 11—15.
- Darlix, J. L. & Rochaix, J. D. 1981 Nucleotide Sequence and Structure of Cytoplasmic 5S RNA and 5.8S RNA of *Chlamydomonas reinhardtii*. *Nucl Acids Res.* 9: 1291—1299.
- Delhas, N., Andersen, J., Sprouse, H.M. & Dudoch, B. 1981 The Nucleotide Sequence of Chloroplast 5S Ribosomal RNA from Spinach. *Nucl. Acids Res.* 9: 2801—2805.
- Erdman, V. A. 1976 Structure and function of 5S and 5.8S RNA. In: Progress in Nucleic acid research and molecular biology, Vol. 18, Academic Press, New York, San Francisco, London 45—90.
- Fox, G. E. & Woese, C. K. 1975 5S RNA Secondary structure. *Nature* 256: 505—507.
- James, M. P. & James, E.M. 1975 Method for predicting RNA secondary structure. *Proc. Nat. Acad. Sci. USA* 72: 2017—2021.
- Jordan, B.R., Gallig, G. & Jourdan, R. 1974 Sequence and Conformation of 5S RNA from *Chlorella* Cytoplasmic Ribosomes: Comparison with other 5S RNA Molecules. *J. Mol. Biol.* 87: 205—225.
- Luehrsens, K. R., Fox, G.E., Kilpatrick, M. W., Walker, R.T., Domadey, J., Krupp, G & Cross, H. J. 1981 The nucleotide Sequence of the 5S rRNA from the archaeobacterium *Thermoplasma acidophilum*. *Nucl. Acids Res.* 9: 965—970.
- Studnicka, G. M., Eiserling, F. A. & Lake, J. A. 1981 A unique secondary folding pattern for 5S RNA corresponds to the lowest energy homologous secondary structure in 17 different prokaryotes. *Nucl. Acids Res.* 9: 1885—1904.
- Tinoco, I., Borer, P. N., Dengler, B. & Levine, M. D. 1973 Improved estimation of secondary structure in Ribonucleic acids. *Nature-New Biology* 246: 40—41.

STUDIES OF 5SRNA SECONDARY STRUCTURE

1. A microcomputer method on the basis of the Fox model

Liu Ciquan, Wang Ying, Zhou Mingpei and Li Jingyan

(*Kunming Institute of Zoology, Academia Sinica*)

This microcomputer method is designed with the Fox model as reference for building 5SRNA secondary structure. By inputting the nucleotide sequences of 5SRNA and calculating through the work of the Z-80 microcomputer about 4 minutes, we got the secondary structure with the lowest total free energy, corresponding to the Fox model when only secondary interactions are under consideration. The result is basically in agreement with the literature (2, 4, 5, 6, 7, 8).

The method has the advantage of simplicity, low cost and consuming less time, so, it is a method that can be easily popularized.

[illegible]

172	-11.8	37	38	39	40	41	42	84	83	82	81	80	79
173	-4.0	38	39	40	96	95	94						
174	-4.0	38	39	40	75	74	73						
175	-6.2	38	39	40	41	96	95	94	93				
176	-6.8	38	39	40	41	42	83	82	81	80	79		
178	-4.0	39	40	41	95	94	93						
179	-4.3	39	40	41	65	64	63						
181	-4.6	39	40	41	42	82	81	80	79				
184	-4.3	40	41	42	104	103	102						
186	-4.3	40	41	42	66	65	64						
188	-9.3	40	41	42	43	78	77	76	75				
190	-4.4	46	47	48	103	102	101						
191	-4.4	46	47	48	80	79	78						
192	-4.4	46	47	48	55	54	53						
193	-6.6	46	47	48	49	103	102	101	100				
194	-6.4	46	47	48	49	87	86	85	84				
195	-4.2	47	48	49	107	106	105						
196	-4.4	47	48	49	102	101	100						
197	-4.4	47	48	49	100	99	98						
200	-6.4	47	48	49	50	85	84	83	82				
201	-4.4	48	49	50	57	56	55						
203	-4.4	50	51	52	89	88	87						
205	-4.2	54	55	56	70	69	68						
206	-4.2	54	55	56	62	61	60						
207	-5.4	59	60	61	65	64	63						
208	-6.4	59	60	61	62	82	81	80	79				
209	-8.2	60	61	62	98	97	96						
210	-4.2	60	61	62	83	82	81						
211	-4.2	60	61	62	81	80	79						
212	-8.2	60	61	62	69	68	67						
214	-9.2	62	63	64	65	76	75	74	73				
215	-9.5	62	63	64	65	66	84	83	82	81	80		
216	-4.2	63	64	65	96	95	94						
217	-4.2	63	64	65	75	74	73						
218	-7.4	63	64	65	66	98	97	96	95				
219	-4.5	63	64	65	66	83	82	81	80				

221 -4.2 64 65 66 97 96 95
 223 -8.2 68 69 70 98 97 96
 225 -4.2 68 69 70 81 80 79
 229 -9.0 73 74 75 76 95 94 93 92
 231 -7.2 74 75 76 94 93 92
 239 -4.2 76 77 78 97 96 95
 247 -21.7 79 80 81 82 83 84 85 86 97 96 95 94 93 92 91 90
 252 -17.2 82 83 84 85 86 94 93 92 91 90
 253 -10.0 83 84 85 93 92 91
 254 -10.0 83 84 85 92 91 90
 256 -15.0 83 84 85 86 93 92 91 90
 257 -9.2 83 84 85 86 91 90 89 88
 258 -10.0 84 85 86 93 92 91
 259 -10.0 84 85 86 92 91 90
 260 -4.2 84 85 86 90 89 88
 264 -6.4 87 88 89 90 108 107 106 105
 265 -6.6 87 88 89 90 101 100 99 98
 266 -4.2 88 89 90 107 106 105
 268 -4.4 88 89 90 100 99 98
 270 -12.2 89 90 91 92 108 107 106 105
 271 -10.0 90 91 92 107 106 105

11 78 144 247

BASE PAIR 1 BOND 119 UA ENERGY =
 BASE PAIR 2 BOND 118 GC ENERGY = -2.200 TOTAL -2.20
 BASE PAIR 3 BOND 117 CG ENERGY = -5.000 TOTAL -7.20
 BASE PAIR 4 BOND 116 CG ENERGY = -5.000 TOTAL -12.20
 BASE PAIR 5 BOND 115 UA ENERGY = -2.200 TOTAL -14.40
 BASE PAIR 6 BOND 114 GC ENERGY = -2.200 TOTAL -16.60
 BASE PAIR 7 BOND 113 GC ENERGY = -5.000 TOTAL -21.60
 BASE PAIR 8 BOND 112 CG ENERGY = -5.000 TOTAL -26.60
 BASE PAIR 9 BOND 111 GU ENERGY = -2.100 TOTAL -28.70
 BASE PAIR 10 BOND 110 GC ENERGY = -2.100 TOTAL -30.80
 INITIAL RING 3.8

BASE PAIR 16 BOND 70 GC ENERGY =
 BASE PAIR 17 BOND 69 CG ENERGY = -5.000 TOTAL -5.00

BASE PAIR	18	BOND	68	GC	ENERGY = -3.200	TOTAL - 8.20
BASE PAIR	19	BOND	67	CG	ENERGY = -5.000	TOTAL - 13.20
INITIAL RING					3.8	
BASE PAIR	31	BOND	51	CG	ENERGY =	
BASE PAIR	32	BOND	50	UA	ENERGY = -2.200	TOTAL - 2.20
BASE PAIR	33	BOND	49	GC	ENERGY = -2.200	TOTAL - 4.40
BASE PAIR	34	BOND	48	AU	ENERGY = -2.200	TOTAL - 6.60
HAIR	35	TO	47		8.0	
BASE PAIR	79	BOND	97	GC	ENERGY =	
BASE PAIR	80	BOND	96	UG	ENERGY = -2.100	TOTAL - 2.10
BASE PAIR	81	BOND	95	GU	ENERGY = - .300	TOTAL - 2.40
BASE PAIR	82	BOND	94	UA	ENERGY = -2.100	TOTAL - 4.50
BASE PAIR	83	BOND	93	GC	ENERGY = -2.200	TOTAL - 6.70
BASE PAIR	84	BOND	92	GC	ENERGY = -5.000	TOTAL - 11.70
BASE PAIR	85	BOND	91	GC	ENERGY = -5.000	TOTAL - 16.70
BASE PAIR	86	BOND	90	GC	ENERGY = -5.000	TOTAL - 21.70
HAIR	87	TO	89		8.0	
TOTAL ENERGEN					SUM = -48.64	